09/479,877

APPLICATION FOR UNITED STATES PATENT

Inventor: Marcia K. Wolf

This application is a continuation of patent application USSN 08/788,145 08/788,145 filed January 24, 1997, which is a continuation-in-part of patent application USSN 243,482 filed May 13, 1994, which is now abandoned.

Field of the Invention:

This invention is related to a CS6 antigen for use in vaccines to protect from pathological effects of enterotoxigenic E. coli.

Background of the Inventions:

CS6 is a component of CFA/IV (colonization factor antigen IV), one of three CFAs commonly found on enterotoxigenic Escherichia coli (ETEC). A recent study showed CS6 on 31% of ETEC isolated from soldiers in the Middle East. Other CFAs and similar proteins found on the surface of ETEC function as adhesins to attach bacteria to intestinal epithelial cells. Attached bacteria can then deliver their toxin(s) to the target cells. It has never been proved that CS6 is an adhesin for human tissue (Knutton, S., M. M. McConnell, B. Rowe, and A. S. McNeish, "Adhesion and Ultrastructural Properties of Human Enterotoxigenic Escherichia coli Producing Colonization Factor Antigens III and IV", Infect.Immun. 57:3364--3371 (1989)), but a study in rabbits indicated CS6 is a colonization factor.

The CS6 operon has much in common with fimbrial operons from E. coli, Salmonella, Yersinia, Klebsiella, Haemophilus, and Bordetella. All contain molecular chaperons and ushers and a

number of structural subunits. This area contains two sequences homologous to insertion sequences, but no complete insertion sequences.

The low GC content (34%) and codon usage that is characteristic of E. coli genes that are expressed at low levels suggest the CS6 genes may have originated in another species. GC content of 35-45% is characteristic of Gram positive bacteria such as Staphylococcus, Streptococcus, Bacillus, and Lactobacillus. Low GC content is common for virulence-associated genes of E. coli.

CS6 is unusual because it is expressed on bacteria grown on a variety of media, unlike other CFA's from ETEC that are only expressed on bacteria grown on CFA agar. This unusual regulation is not peculiar to strain E8775 because ETEC isolated in 1990 expressed CS6 when grown on L agar. Temperature regulation of CS6 expression is characteristic of other CFA's from ETEC and virulence genes in a variety of pathogenic bacteria.

Although CS6 has never been visualized by negative staining, electron microscopy using anti-CS6 sera and colloidal gold indicated that it is present on the surface of ETEC. The apparent major protein associated with CS6 is approximately 16 kDa which is in the range of molecular weights typical for subunits for fimbriae and fibrillae. CS6 from ETEC strain E10703 of serotype O167:H5 has been cloned (Willshaw, et al., FEMS Microbio. Let. **49:** 473-478 (1988)). Only 3 kb of DNA was necessary for expression of CS6. That is in contrast to fimbriae that typically require approximately 9 kb of DNA and include genes for subunits as well as proteins

for transport of subunits and synthesis and assembly on the bacterial surface.

Grewal teaches bacterial strains transformed with plasmids containing genes which encode CS6. However, that reference does not teach use of plasmids under the controls of a lac promotor and a CS6 promotor.

Brief Description of the Drawings:

Figure 1 shows the restriction sites and the location of the pertinent genes that make up the CS6 operon.

Figure 2 shows derivation of the clone containing the kanamycin resistance gene.

Description of the Invention:

It is the purpose of this invention to provide structural proteins which will act as antigens to stimulate protective antibodies against enterotoxigenic Escherichia coli. Particularly important are proteins having the antigenic properties of the proteins encoded by the cssA and cssB genes. Constructs may be prepared which encode either one or both of the proteins. However, both proteins would be needed to provide desirable protection. The CS6 operon includes four genes which we designate cssA, cssB, cssC, and cssD. cssA and CssB encode the structural proteins of CS6. The CS6 operon has much in common with fimbrial operons from E. coli, Salmonella, Yersinia, Klebsiella, Haemophilus, and Bordetella. All contain molecular chaperons and ushers and a number of structural subunits. In a preferred embodiment, plasmids containing all four genes are transformed into attenuated bacteria, which are

then given by mouth to prevent morbidity arising from infection with E. coli.

CS6 has two major subunits; protein sequencing data demonstrates that CS6A and CS6B are both present. The DNA sequence yields a mechanism for expression of similar amounts of the two proteins. The CS6 operon contains DNA immediately downstream of cssB which can form a stem-loop with a stem rich in G and C which commonly act as transcription terminators. Termination at this site yields a transcript with cssA and cssB such that CssA and CssB proteins would be translated in equal amounts. Fimbrial operons for Pap, K99, and K88 have stem loops immediately downstream of the genes for the major coding structural subunits. This has been offered as a mechanism for over expression of subunit genes relative to other genes in the operons. In the case of CS6, this would allow over expression of both CS6A and CS6B.

The occurrence of two major structural proteins is unusual because fimbriae have a single major subunit and a number of minor subunits. CS3, which has been designated fibrillar rather than fimbrial, is an exception to this generality because it has 2 subunits. CssD belongs to the family of molecular ushers located in the outer membrane that accept subunits from the chaperone and escorts them to the bacterial surface. Apparently the entire cssD gene is not necessary for CS6 expression since CS6 was detected from clones carrying pDEP5 which only contains the N-terminal one-third of cssD. Klemm and Christiansen found that mutations in the usher for Type 1 fimbriae reduced fimbriation but 10% of the

bacteria produced a few fimbriae (Mol.Gen.Genet. 220:334-338).

The CS6 proteins are produced in the transformed bacteria and are present on the exterior surface of the bacteria. These proteins give rise to immunological response in the host. For immunization, the bacteria may be given either dead or alive. When attenuated bacteria have been transformed, the bacteria can be given live in mildly basic carriers. Economical and readily available carriers include carbonated water which may be flavored. The administration of the transformed bacteria in carbonated beverages is particularly useful, since the means necessary for administration is widely available.

In a preferred embodiment, the products are produced under the control of a lac promotor from pUC19. In the preferred embodiment, a vector pM346 containing a kanamycin resistant gene makes it possible to provide products which are appropriate for use in humans.

The CS6 proteins may also be extracted from the supernatant of the culture containing the organisms which express the proteins. The proteins may then be administered orally. The proteins may be formulated by means known in the art, including microencapsulation, coated capsules and liposomes. The proteins may be lyophilized before formulation.

MATERIALS AND METHODS

Source of nucleic acid

The genes for CS6 expression were from enterotoxigenic Escherichia coli (ETEC) strain E8775 tox⁻ of serotype O25:H42 which

was a gift from Alejandro Cravioto. E8775 tox⁻ is a derivative of E. coli strain E8775 which was originally isolated from Bangladesh. DH5α which was purchased from Bethesda Research Laboratories, Inc., Gaithersburg, MD. pUC19 was originally purchased from P-L Biochem.

The antibiotic resistance gene encodes resistance to kanamycin and was purchased from Pharmacia, Uppsala, Sweden (Kan^R GenBlock®).

CS6 expression is regulated from its native promoter. That is demonstrated by retention of control by growth temperature and is consistent with the DNA sequence determined from the clone. A contribution of the lac promoter from pUC19 is undefined. The contribution of increased copy number of the plasmid is probably substantial.

The nucleotide sequence containing the coding region was determined to be as constructed containing the kanamycin resistance gene was as follows:

```
  1  AAGCTTGTAA CCAGTTGATA AAAATATATC ACGCTGGGAA TGACGTGATG
 51  TATATACGGA GCAGCTATGT CGGAACAGAT ATTTTCCTAT CGGTATGCGT
101  TGTGAGTAAG CGTAAAGCCA ATGCTGTCTG TAACTCCTGA TCCTTGCAGA
151  CTAAATTAGA GCTCCTTCTA AATTAGACGG ATGGATAAAC CTACAGACTG
201  GCGCTCTGGG TCTCGCCGGA TATTTTCTAA TGAATTTAAG CTTCATATGG
251  TTGAACTGGC TTCGAAACCA AATGCCAATG TCGCACAACT GGCTCGGGAA
301  CATGGCGTTG ATAACAACCT GATTTTTAAA TAGCTACGCC TCTGGCAAAG
351  AGAAGGACGT ATTTCTCGTA GAATGCCTCC AACTATTGTA GGCCCTACAG
401  TACCACTGAG GTAGCCTGAA TTTAAAGCCG AAGCGGTCAG AACTGTTCTT
451  GGTGTGAACG TAGCACTCAC CAATAAAAGC ATCAATACGG TGCTCTGTTG
501  ACACATTACG AATGTTATGT ATACAATAAA AATGATTATA GCAATATTAA
```

551 TGGTGTTATA TGAAGAAAAC AATTGGTTTA ATTCTAATTC TTGCTTCATT

601 CGGCAGCCAT GCCAGAACAG AAATAGCGAC TAAAAACTTC CCAGTATCAA

651 CGACTATTTC AAAAAGTTTT TTTGCACCTG AACCACGAAT ACAGCCTTCT

701 TTTGGTGAAA ATGTTGGAAA GGAAGGAGCT TTATTATTTA GTGTGAACTT

751 AACTGTTCCT GAAAATGTAT CCCAGGTAAC GGTCTACCCT GTTTATGATG

801 AAGATTATGG GTTAGGACGA CTAGTAAATA CCGCTGATGC TTCCCAATCA

851 ATAATCTACC AGATTGTTGA TGAGAAAGGG AAAAAAATGT TAAAAGATCA

901 TGGTGCAGAG GTTACACCTA ATCAACAAAT AACTTTTAAA GCGCTGAATT

951 ATACTAGCGG GGAAAAAAAA ATATCTCCTG GAATATATAA CGATCAGGTT

1001 ATGGTTGGTT ACTATGTAAA CTAAATACTG GAAGTATGAT TATGTTGAAA

1051 AAAATTATTT CGGCTATTGC ATTAATTGCA GGAACTTCCG GAGTGGTAAA

1101 TGCAGGAAAC TGGCAATATA AATCTCTGGA TGTAAATGTA AATATTGAGC

1151 AAAATTTTAT TCCAGATATT GATTCCGCTG TTCGTATAAT ACCTGTTAAT

1201 TACGATTCGG ACCCGAAACT GGATTCACAG TTATATACGG TTGAGATGAC

1251 GATCCCTGCA GGTGTAAGCG CAGTTAAAAT CGCACCAACA GATAGTCTGA

1301 CATCTTCTGG ACAGCAGATC GGAAAGCTGG TTAATGTAAA CAATCCAGAT

1351 CAAAATATGA ATTATTATAT CAGAAAGGAT TCTGGCGCTG GTAACTTTAT

1401 GGCAGGACAA AAAGGATCCT TTCCTGTCAA AGAGAATACG TCATACACAT

1451 TCTCAGCAAT TTATACTGGT GGCGAATACC CTAATAGCGG ATATTCGTCT

1501 GGTACTTATG CAGGAAATTT GACTGTATCA TTTTACAGCA ATTAAAAAAA

1551 GGCCGCATTA TTGCGGCCAT TGACGATACT GCTAGGCAAA AATATGAAAT

1601 CAAAGTTAAT TATACTATTG ACGTTAGTGC CATTTTCATC TTTTTCAACA

1651 GGAAATAATT TTGAAATAAA TAAGACACGA GTAATTTACT CTGACAGCAC

1701 ACCATCAGTT CAAATATCAA ATAATAAAGC ATATCCTTTA ATTATTCAAA

1751 GCAATGTATG GGATGAAAGC AATAATAAAA ATCATGACTT TATAGCAACA

1801 CCACCGATTT TTAAAATGGA AAGTGAAAGT CGGAATATAA TAAAAATAAT

1851 TAAAACAACT ATTAATTTGC CGGACTCTCA GGAAAGTATG AGATGGTTAT

1901 GTATTGAATC AATGCCACCA ATAGAAAAAA GTACTAAAAT AAACAGAAAA

1951 GAAGGAAGGA CAGACAGTAT TAATATCAGC ATTCGGGGGT GCATTAAACT

2001 GATATATCGA CCTGCCAGTG TTCCGTCTCC TGTTTTTAAT AATATAGTAG

2051 AAAAATTAAA ATGGCATAAA AATGGAAAGT ATCTTGTATT AAAAAATAAT

2101 ACACCCTATT ACATTAGCTT TTCTGAGGTT TTTTTTGATT CAGATAAAGT

2151 AAACAATGCA AAAGATATTT TATATGTAAA ACCATACTCA GAGAAGAAAA

2201 TAGATATCAG CAACAGAATA ATAAAAAAAA TCAAATGGGC TATGATTGAT

2251 GATGCTGGCG CAAAAACAAA ACTTTATGAA TCAATTTTAT AAAAAATCTC

2301 ATTACAGTAT ACAAAAACAT CAGATTACAG GCTTGCTTTT TTTGCTATTT

2351 ATATATCCTT TCTCAACCTC ATATGGAAAT GAACAATTTA GTTTTGACTC

2401 ACGATTCCTA CCATCAGGTT ATAATTACTC TTTAAATAGT AACTTACCTC

2451 CTGAAGGTGA GTATCTGGTT GATATTTATA TTAACAAAAT AAAAAAGGAG

2501 TCCGCGATTA TTCCTTTTTA TATAAAAGGA AATAAACTTG TACCATGTTT

2551 ATCAAAAGAA AAAATTTCAT CTTTGGGTAT CAACATTAAT AATAACGACA

2601 ACACAGAGTG TGTAGAAACA AGTAAGGCAG GTATTAGTAA TATCAGCTTT

2651 GAGTTTAGCT CTCTTCGTTT GTTTATTGCT GTACCGAAAA ATCTTCTGTC

2701 TGAGATTGAT AAAATATCAT CAAAGGATAT AGATAACGGG ATTCATGCTT

2751 TATTTTTTAA TTATCAAGTA AATACAAGGC TAGCCAATAA TAAAAATCGT

2801 TATGATTACA TTTCTGTTTC ACCAAATATA AATTATTTTT CATGGCGGTT

2851 GCGTAATCTT TTTGAATTTA ACCAAAACAA CGATGAAAAA ACATGGGAAA

2901 GAAACTACAC TTATCTAGAA AAAAGTTTTT ATGATAAAAA GCTAAACTTA

2951 GTCGTTGGTG AAAGTTATAC GAATTCAAAT GTTTATAATA ACTACTCTTT

3001 TACTGGTATT TCAGTTTCTA CAGATACAGA TATGTATACG CCAAGTGAAA

3051 TCGATTATAC ACCAGAAATT CATGGAGTGG CTGATTCAGA CTCTCAGATT

3101 ATTGTCAGGC AAGGCAACAC CATTATCATT AATGAAAGTG TTCCAGCCGG

3151 ACCGTTCTCA TTTCCAATAA CCAATCTCAT GTATACTGGG GGGCAACTTA

3201 ATGTGGAGAT AACAGATATT TATGGAAATA AAAAACAATA TACTGTCAAT

3251 AATTCCTCTC TTCCTGTTAT GAGAAAAGCG GGACTAATGG TATATAATTT

3301 TATATCTGGG AAATTAACAA AAAAAAATAG TGAGGATGGT GATTTTTTTA

3351 CTCAAGGTGA TATTAACTAC GGTACTCACT ATAACAGCAC ACTATTCGGT

3401 GGATATCAGT TTAGTAAAAA TTATTTTAAC TTATCTACTG GTATAGGCAC

3451 TGATCTGGGA TTTTCTGGAG CATGGCTACT ACACGTTAGC AGAAGTAATT

3501 TTAAGAATAA AAATGGATAT AATATTAATC TACAACAAAA CACTCAGTTA

3551 AGACCATTCA ATGCCGGGGT TAATTTCGAT TACGCATACA GAAAAAAAAG

3601 GTATGTGGAA CTTTCCGACA TTGGCTGGCA TGGTAATTTA TATAATCAAC

3651 TTAAAAATAG TTTTTCTTTA TCCTTGTCAA AATCATTGAA TAAATACGGA

3701 AATTTCTCAC TTGATTATAA CAAAATGAAA TACTGGGATA ATGCGTATGA

3751 TAGTAACTCA ATGTCGATTC GTTATTTTTT TAAATTCATG CGAGCAATGA

3801 TTACAACAAA TTGTTCTTTA AATAAATATC AATCTTATGA AAAAAAAGAT

3851 AAAAGATTTA GTATTAATAT ATCATTGCCT TTAACCAAAG ATTACGGGCA

3901 CATATCTTCA AACTATTCAT TTTCCAATGC AAATACAGGA ACGGCAACCA

3951 GTTCTGTAGG CTTAAACGGT AGTTTTTTTA ATGACGCAAG ATTAAACTGG

4001 AACATTCAGC AGAACAGAAC GACCCGTAAC AATGGATATA CTGATAATAC

4051 CAGTTACATA GCAACCAGCT ATGCCTCTCC CTATGGCGTT TTTACTGGTT

4101 CATATTCAGG ATCGAACAAG TATTCAAGCC AGTTTTATTC TGCATCGGGA

4151 GGTATTGTTT TGCATAGCGA TGGCGTAGCT TTTACTCAAA AAGCCGGAGA

4201 TACCTCTGCT CTTGTCCGTA TTGATAATAT TTCTGATATA AAAATTGGTA

4251 ACACTCCTGG TGTTTATACT GGGTATAATG GTTTTGCTTT AATTCCTCAT

4301 CTTCAGCCGT TCAAAAAAAA CACCATTTTA ATTAATGATA AAGGAATTCC

4351 AGACGGTATT ACTCTTGCTA ATATAAAAAA ACAAGTTATC CCATCACGAG

4401 GAGCTATTGT TAAAGTAAAA TTTGATGCTA AAAAAGGCAA TGACATTTTG

4451 TTTAAGCTTA CAACTAAAGA TGGAAAAACG CCCCCATTAG GAGCTATAGC

4501 CCATGAAAAA AATGGAAAAC AGATTAATAC GGGTATCGTT GACGATGATG

4551 GTATGCTTTA TATGTCTGGA TTATCAGGGA CAGGGATTAT TAATGTAACA

4601 TGGAATGGAA AAGTCTGTTC ATTTCCTTTT TCAGAAAAAG ATATATCTAG

4651 CAAACAATTA TCTGTTGTAA ATAAACAATG TTAGGTAGTG CATCCAATTA

4701 GTAGAACATG TGTTTTTCGA TAAACGCTCC GATCTCTTTT TCGTGGATCT

4751 CAACTGAGCG TGAGAAGCAG ATTGTTTTAC GAGCCAACCG CTTAATGCGG

4801 GTGCGTAGCG TCAGATTATT ACGCTCAATG CGTTGGGTGA ATATTTTGCC

4851 GGTCAGATGC TTATTCTTCG GTACC **Sequence ID No. 1**

B. Cell expression clone:

E. coli HB101 was purchased from the American Type Culture Collection, Rockville, Maryland. It is ATCC #33694 and batch #91-1. (Escherichia coli ATCC 33694)

Preceptrol [Reg TM] culture. D. Ish-Horowicz and J.F. Burke HB101 <--- H. Boyer. Genotype: F- leuB6 proA2 recA13 thi-1 ara-14 lacY1galK2 xyl-5 mtl-1 rpsL20 supE44 hsdS20 (r- B m- B at least thi-hsd from Escherichia coli B). Produces isoprene (Curr. Microbiol. 30:97-103, 1995). J. Mol. Biol. 41: 459-472, 1969; Methods Enzymol. 68: 245-267, 1979.) Growth Conditions: Medium 1065 37C.

The plasmid containing the CS6 genes, the pUC19 origin of replication, and the gene for kanamycin resistance was transferred into E. coli HB101 by transformation. Transformants were selected by growth on L agar supplemented with 0.04% Xgal with 50 μgm per ml kanamycin sulfate and/or 50 μgm per ml ampicillin.

One copy of the CS6 genes exists as an extrachromosomal

plasmid of high (500-700) copy number. The CS6 genes are present on a plasmid, not integrated into the chromosome. The plasmid has been isolated from the strain and examined by agarose gel electrophoresis.

Plasmid DNA from E8775 tox$^-$ was transferred to laboratory strain DH5α as a cointegrate with F' lac_{ts}::Tn5, a conjugative plasmid. Transfer of the F' lac_{ts}::Tn5 plasmid was selected by antibiotic resistance to kanamycin and CS6 expression was detected by Western blot using polyclonal antisera specific for CS6. Plasmids were isolated and a cointegrate was identified based on the large size. A spontaneous derivative in which the F' lac_{ts}::Tn5 was removed was obtained and named M56. M56 contains a 61-megadalton plasmid from E8775 tox$^-$ and expresses CS6. Plasmid DNA from M56 was isolated, partially digested with restriction enzyme HindIII, and ligated to pUC19 that had been digested with HindIII. The ligation mixture was transformed into DH5α and plated onto L agar supplemented with ampicillin and X-gal. White (lac$^-$) colonies were picked to CFA plates supplemented with ampicillin and tested for CS6 expression.

A stable clone named M233 with an insert of approximately 24 kb into the cloning site of pUC19 was obtained. It was a spontaneous deletion of a larger clone. Subclones were obtained by digestion with various enzymes and a subclone containing approximately 5 kb from the HindIII site to KpnI site was found that expressed CS6. This clone was designated M285. Expression of CS6 was verified by transferring plasmids into E. coli strain HB101 and

detecting CS6 expression. The cloned CS6 is expressed under the same conditions as CS6 from the native 61-megadalton plasmid: CS6 was detected in extracts from bacteria grown at 37°C on CFA agar, L agar or MacConkey agar. CS6 was not expressed on bacteria grown at 17°C.

Studies were performed to determine appropriate handling of strain M285 for reproducible expression of CS6. Growth temperature was found to be especially important.

As indicated above, the protein sequence of the N-terminus of CS6 was determined from strains E8775 and from M233, the large clone derived from E8775. The 16 kDa proteins recovered from heat, saline extracts, and ammonium sulfate precipitation of M233 yielded two amino acids at each position (except cycle 12) indicating that two proteins were present. From the strength of the two signals, a probable primary sequence and a probable secondary sequence call was made for each of fifteen cycles. Quantitative analysis of the peak areas indicated that the molar ratio of the primary sequence (CS6A) to secondary sequence (CS6B) was approximately 3:1. The presence of the same two proteins was evident from strain E8775 grown on CFA agar and on L agar.

The DNA sequence of the DNA inserted into pUC19 in clone M285 was determined. Wim Gaastra's group in the Netherlands independently determined the DNA sequence of CS6 genes from ETEC strain E10703. The DNA sequences are available from Genebank accession numbers U04846 and U04844, respectively. A stretch of DNA of 4,219 base pairs was 98% identical. The DNA sequences diverge abruptly

on both sides of the common region, defining the limits of the CS6 genes. Four open reading frames were detected within the common area. These were designated cssA, cssB, cssC, and cssD.

The four open reading frames are preceded by consensus sequences for binding RNA polymerase and ribosomes. The first open reading frame, cssA was identified as the gene for the CS6 structural protein designated as the primary protein based on the amino acid N-terminal sequence. The deduced molecular weight agrees with that previously determined from SDS PAGE. cssA includes a signal sequence that is typical for many exported proteins. Eleven of 136 residues differ between the deduced CssA proteins from E8775 and from E10704.

cssB begins 17 bases downstream from cssA. There is a typical signal sequence. cssB was identified as the gene for the CS6 structural protein designated as the secondary protein based on the amino acid N-terminal sequence. Five of 146 residues differ between the deduced CssA proteins from E8775 and from E10704.

A region of dyad symmetry is present 6 bases downstream from CssB in both clones. The sequence is GGCCGCATTATTGCGGCC (**Sequence #2**) in E8775 ETEC and GGCCGCATTATTGATTGCGGCC (**Sequence #3**) in E10703. Underlined bases form the G-C rich stem. The calculated free energy value of these structures is -14.8 kcal. Such structures are often found in fimbrial operons after the genes encoding structural proteins.

cssC begins 48 bases downstream from cssB. It has a typical signal sequence. The deduced proteins from both clones have 212

residues with 7 differences. A search of protein databases indicated CssC is homologous to chaperone proteins necessary for expression of a number of fimbriae. The structure of PapD, the chaperone protein for Pap fimbriae, has been solved by X-ray crystallography and regions important for conserving the structural domains have been identified. CssC conforms to this consensus.

The cssD gene begins 14 bases upstream of the end of cssC. The protein from E8775 is truncated relative to the protein from E10703 and there are 28 differences between CssD from E8775 and E10703. The deduced protein from cssD is homologous to molecular ushers. Overall, CssD and the other proteins are only around 30% identical and around 50% similar, but the nine proteins have areas of high homology dispersed throughout, especially the first 410 residues, and 4 cysteines (residues 91, 112, and two near the C-terminus) which are conserved in all ushers.

A region of dyad symmetry is present 347 base pairs into the CssD gene in both clones. The calculated free energy value of these structures is -7.2 kcal.

The plasmid from strain M285 was transformed into E. coli HB101 purchased from ATCC. The resulting strain was named M295. Expression of CS6 from M295 was achieved from small-scale fermentation. For production for human use, it was desirable to add a gene for resistance to kanamycin as the selectable marker. To that end, a vector was constructed based on pUC19 but with a gene for kanamycin resistance in place of the gene for ampicillin resistance. The CS6 genes from the pUC19 clone were subcloned into the

new vector and transformed into E. coli HB101.

Vector pM323 was constructed as follows. The kanamycin resistance gene was purchased from Pharmacia, Uppsala, Sweden (Kan^R GenBlock®) and inserted into a cloning vector by Dr. David Lanar at WRAIR. DNA including the gene was amplified by PCR using the plasmid from Dr. Lanar as template and primers flanking the multiple cloning site. A product of the desired size (1,580 bp) was obtained, but with much template present. To increase the purity of the 1,580 fragment, a second PCR reaction was performed, this time with a small amount of the first PCR reaction as template. This product was confirmed by agarose gel electrophoresis, then digested with restriction enzyme HincII to remove unwanted restriction enzyme recognition sites. This product was ligated to pUC19 digested with SspI. The ligation mix was transformed into E. coli DH5α and plated on L agar plates supplemented with kanamycin and Xgal. Isolate M318 had the desired phenotype of resistance to kanamycin and ampicillin with lacZ' intact. The gene for ampicillin resistance was removed to make a smaller vector. This was achieved by designing and synthesizing 2 oligonucleotides to amplify just the portion of pM318 with the gene for kanamycin resistance, the lacZ' gene carrying the multiple cloning site, and the origin of replication. PCR was performed, the product ligated then transformed into E. coli DH5α with selection on L agar plates supplemented with kanamycin and Xgal. Isolate M323 had the desired phenotype of resistance to kanamycin, sensitivity to ampicillin, and intact lacZ'. Restriction digest

patterns confirmed the plasmid was a derivative of pUC19 with the gene for kanamycin resistance in place of the gene for ampicillin resistance.

The CS6 genes were cloned into vector pM323 from pM285. pM323 and pM285 were digested with restriction enzyme SstI, ligated, and transformed into E. coli DH5α with selection on L agar plates supplemented with kanamycin and Xgal. Isolate M334 was determined to express CS6. Plasmid analysis revealed M334 carried the CS6 genes and 2 copies of the vector. An attempt was made to remove one copy of vector and at the same time move the clone into HB101, the desired host strain for fermentation. Isolate M340 was determined to express CS6 and retained 2 copies of the vector. An isolated colony of M340 was shown to produce high amounts of CS6 and was saved as M346.

In another embodiment lacking the kanamycin resistance gene, clones from an ETEC strain of serotype O25:H42 were derived from E. coli E8775 which was originally isolated from samples from Bangladesh. E. coli M56, which contains a 61-megadalton plasmid from E8775 Tox$^-$ and expresses CS6 has been described. The host for cloning was E. coli DH5α which was purchased from Bethesda Research Laboratories, Inc., Gaithersburg, MD. The host for plasmids used for production of heat, saline extracts was HB101 (EMBO J. 4:3887-3893 (1985)).

Clones from E8775 were routinely grown in L broth. Antibiotics were added to L broth supplemented with agar as follows. Ampicillin was added, when appropriate, at 50 μg/ml. Chloramphenicol

col was used at 30 μg/ml. X-Gal (5-bromo-4-chloro-3-indolyl ß-D-galactopyranoside, Sigma) was added at 0.004%. CFA plates were prepared as previously described (Infect.Immun. **57**:164-173 (1989)).

Cloning CS6 from E8775. The 61-megadalton plasmid from E. coli M56 was partially digested with HindIII and ligated to pUC19 that had been digested with HindIII. The ligation mixture was transformed into E. coli DH5α and plated onto L agar plates supplemented with ampicillin and X-gal. White (lac$^-$) colonies were picked to CFA plates supplemented with ampicillin and tested for CS6 expression using antisera as described below. Plasmids were purified as described (Infect.Immun. **57**:164-173 (1989)). Restriction enzymes were used according to the manufacturer's directions.

Detection of CS6 Expression. CS6 expression by bacterial colonies was detected after transfer to nitrocellulose and treatment as described by Mierendorf (Methods Enzymol. **152**:458-469 (1987)). Primary antisera was specific for CS6 and was raised in rabbits and absorbed as previously described (Infect.Immun. **57**:164-173 (1989)), except that rabbits were inoculated intravenously with live bacteria suspended in normal saline. Secondary antibody was peroxidase-conjugated goat anti-rabbit IgG (Cappel Laboratories, Cochranville PA) and detection was by TMB Substrate (Kirkegaard & Perry Laboratories, Inc., Gaithersburg MD).

Positive identification of CS6 was by western blots of heat, saline extracts. Heat, saline extracts were prepared from bacteria grown on the indicated media as described (Infect.Immun. **27**:657-666 (1980)). Proteins were separated on precast 16% Tricine sodium

dodecyl sulfate-polyacrylamide gels (SDS-PAGE, Novex Novel Experimental Technology, San Diego, CA) and transferred to nitrocellulose. Blots were handled as described above for colony blots.

Determination of N-terminal sequence. Heat, saline extracts were obtained from E8775 or clones of E8775 grown on L agar or CFA. Partial purification of CS6 was obtained by ammonium sulfate precipitation, with extracts sequentially precipitated at 20%, 40%, then 60% saturation. Samples at 40% and 60% saturation were dialyzed against deionized water and loaded onto precast 16% Tricine SDS-PAGE (Novex, San Diego, CA). Proteins were blotted onto polyvinylidene difluoride (PVDF) membranes (Westrans, Schleicher & Schuell, Keene, NH), stained by Coomassie blue (Rapid Coomassie Stain, Diversified Biotech, Newton, MA) and bands of approximately 16 kDa were excised for automated gas-phase N-terminal sequencing analysis (Applied Biosystems Model 470A, Foster City, CA). Data were analyzed using the Model 610A Data Analysis Program, Version 1.2.1 (Applied Biosystems, Inc, Foster City, CA). These methods have been described in detail (Infect.Immun. 60:2174-2181 (1992)).

DNA sequencing. DNA sequencing of the clones derived from E8775 was performed using the Model 373A DNA sequencing system from Applied Biosystems, Inc, Foster City, CA. Reactions were performed using the dideoxy method with fluorescent dye-labeled terminators, double-stranded templates, oligonucleotide primers, and AmpliTaq DNA polymerase following the manufacturer's protocol. Appropriate oligonucleotide primers were synthesized using a Model 391 DNA

Synthesizer (Applied Biosystems, Inc, Foster City, CA). Plasmids were purified for use as templates by a slight modification of the alkaline lysis method and cesium chloride density gradient centrifugation described by Maniatis (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor (1982)). Plasmids were recovered by dialysis followed by multiple ethanol precipitations to remove residual salt. Sequence analysis was performed using software developed by the University of Wisconsin Genetics Computer Group (Nucleic Acids Res. **12**:387-395 (1984)).

RESULTS

CS6 genes cloned from ETEC strain E8775 into pUC19. A stable clone named M233 was obtained from a partial digest of the 61-megadalton plasmid from E. coli M56. It was a spontaneous deletion of a larger clone. The insert in M233 was approximately 24 kb. Subclones were obtained by digestion with various enzymes and a subclone containing 4.9 kb from the HindIII site to KpnI was found that expressed CS6. This clone was designated M285. Expression of CS6 was verified by transferring plasmids into E. coli HB101 and detecting CS6 in heat, saline extracts. The cloned CS6 is expressed under the same conditions as CS6 from the native 61-megadalton plasmid (Table 1). CS6 was detected in western blots of heat, saline extracts of bacteria grown on CFA, L agar or MacConkey agar. CS6 was not expressed on bacteria grown at 17°C.

Table 1. Regulation of CS6 Expression

Strain:	M287	M56	E8775	HB101
Plasmid:	pM285	native	native	none
Chromosome:	HB101	HB101	native	HB101
Media	-----	------	------	-----
CFA 37°C	+	+	+	-
CFA 17°C	-	-	-	-
L agar	+	+	+	-
MacConkey	+	+	+	-

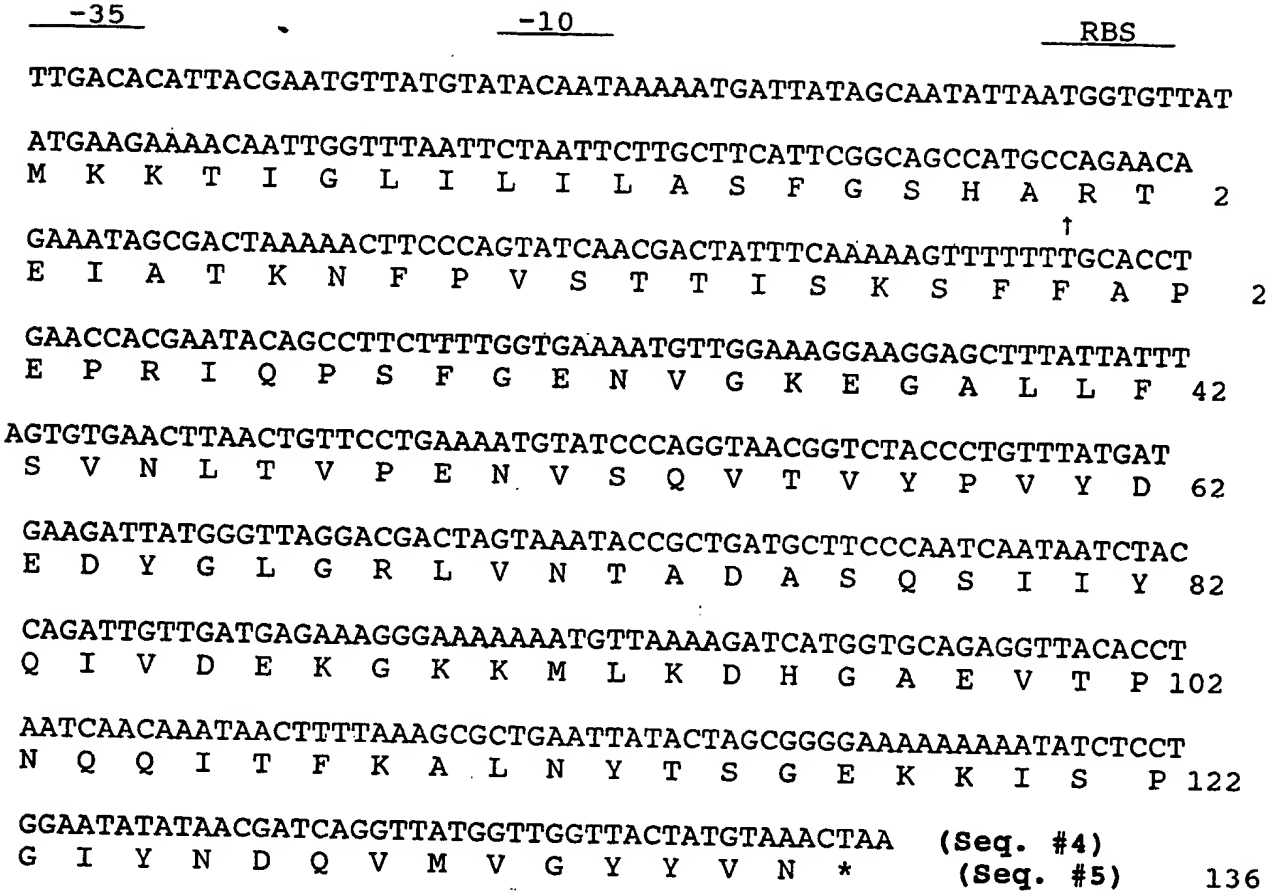
N-terminal sequence of CS6. The protein sequence of the N-terminus of CS6 was determined from strains E8775 and from M233, the large clone derived from E8775. The 16 kDa proteins recovered from heat, saline extracts, and ammonium sulfate precipitation of M233 yielded two amino acids at each position (except cycle 12) indicating that two proteins were present. From the strength of the two signals, a probable primary sequence and a probable secondary sequence call was made for each of fifteen cycles. Quantitative analysis of the peak areas indicated that the molar ratio of the primary sequence (CS6A) to secondary sequence (CS6B) was approximately 3:1. The presence of the same two proteins was evident from strain E8775 grown on CFA agar and on L agar.

DNA sequence of CS6 operons. The sequences of DNA cloned from E8775 (in M285) were determined. They are available from Genebank accession number U04846. The DNA sequence, when compared with sequences from another strain, were found to diverge abruptly on

both sides of the common area. Four open reading frames were detected. These were designated cssA, cssB, cssC, and cssD for CS six.

The GC content of the DNA is 34% and the codon usage is in the range found for Escherichia coli genes that are expressed at low or very low levels as defined by Osawa et al (Prokaryotic Genetic Code. Experentia **46**:1097-1106 (1990)).

Genes encoding CS6 structural genes. The four open reading frames are preceded by consensus sequences for binding RNA polymerase and ribosomes. DNA and deduced amino acid sequence of cssA, a CS6 structural protein. The DNA sequence of the entire operon is available from Genebank accession number U04844. The deduced amino acid sequence from E8775 is given. The arrow indicates the site of cleavage of the signal peptide. The protein sequence is associated with the sequence for the second construct:



The first open reading frame, cssA was identified as the gene for the CS6 structural protein CS6A designated as the primary protein based on the amino acid N-terminal sequence. cssA includes a signal sequence that is typical for many exported proteins. The deduced CssA protein from E8775 has 136 residues, as shown above and in Table 2. The molecular weight agrees with that previously determined from SDS PAGE. No homologous proteins were found by searching the protein databases, but conserved residues are present near the C-terminus and this is typical of fimbrial subunits that are carried across the periplasm by chaperons.

Table 2. Characteristics of Proteins Deduced from CS6 Operons

Protein	Source	Number of Residues	Molecular Weight	Isoelectric Point
CssA (CS6A)	E8775	136	15,058	5.27
CssB (CS6B)	E8775	146	15,877	4.40
CssC	E8775	212	24,551	10.24
CssD	E8775	802	90,393	9.97

cssB begins 17 bases downstream from cssA. There is a typical signal sequence. cssB was identified as the gene for the CS6 structural protein CS6B designated as the secondary protein based on the amino acid N-terminal sequence. The C-terminus matches the consensus typical of fimbrial subunits. The sequence from E8775 is given. The arrow indicates the site of cleavage of the signal peptide.

MLKKIISAIA LIAGTSGVVN A↑GNWQYKSLDV NVNIEQNFIP DIDSAVRIIP 30
VNYDSDPKLD SQLYTVEMTI PAGVSAVKIA PTDSLTSSGQ QIGKLVNVNN 80
PDQNMNYYIR KDSGAGNFMA GQKGSFPVKE NTSYTFSAIY TGGEYPNSGY 130
SSGTYAGNLT VSFYSN 146 **(Seq. #6)**

A region of dyad symmetry is present 6 bases downstream from CssB in both clones. The sequence is GGCCGCATTATTGCGGCC **(Seq. #2)** in E8775 ETEC. Underlined bases form the GC rich stem.

Genes with homology to fimbrial accessory proteins. cssC begins 48 bases downstream from cssB. It has a typical signal sequence. The deduced proteins from both clones have 212 residues with 7 differences. A search of protein databases indicated CssC is homologous to chaperone proteins necessary for expression of

Pap, CS3, K88, K99, CS31A, S, and Type 1 fimbriae of E. coli and SEF14 of Salmonella enteritidis, F1 and pH6 antigen of Yersinia pestis, Type 3 of Klebsiella pneumoniae, Type b of Haemophilus influenzae, and filamentous heamagglutinin of Bordetella pertussis. The structure of PapD, the chaperone protein for Pap fimbriae, has been solved by X-ray crystallography and regions important for conserving the structural domains have been identified. CssC conforms to the following consensus. Below is the deduced amino acid sequence of cssC. The * indicates conservative amino acid replacements. Dots are gaps necessary for aligning all sequences. Boxes indicate beta strands as defined for PapD. The designation of the beta strands for domain 1 (A1 through G1) and domain 2 (A2-G2) are given below each box.

*	* *R***		* *		*** ** *		****P*****
NNF	EINKTRVIYS	DSTP	SVQISNN	KAYP..	LIIQSNVWDES	NNKNH..D	FIATPPIFKM
	A1		B1		C1		D1

*	* ****	** E	** ** * *	P* * * *	
ESES	RNIIKIIK	TTI..NLPDSQE	SMRWLCIESM	PPIEKST..KINRKEGRTDSINISI	110
	E1		F1		

*K*****	P* *	*	*	* * N	*P**	*** *	
RGCIKLIYR	PASVPSPVFNN.IVEK	LKWHK	NGKY	LVLKN	NTPYY	ISFSEVF	160
G1		A2		B2		C2	

*	* *	P	* *	* *	*	*D G*	
FDSDKV..NNAKD	ILYVK	PY	SEKKID	ISN..RIIKKI	KWAMI	DDAGAKT	KLYESIL
	D2		E2		F2		G2

(Seq. #7)

CssD begins 14 bases upstream of the end of cssC. When compared with a second sequence there are 28 differences between CssD from E8775 and the other sequence. The deduced protein from cssD is homologous to molecular ushers found in operons of Pap, CS3, K88, K99 and Type 1 fimbriae of E. coli and SEF14 of Salmonella enteritidis, F1 of Yersinia pestis, and Type 3 of Klebsiella pneumoniae. Overall, CssD and the other proteins are only around 30% identical and around 50% similar. Asterisks above the CssD sequences indicates amino acids conserved relative to molecular ushers.

```
                    *****  *  **          *    **
MNQFYKKSHYSIQKHQITGLLFLLFIYPFSTSYGNEQFSFDSRFLPSGYN   50

    *  ** G Y *** *N          * *        * C**   *  *
YSLNSNLPPEGEYLVDIYINKIKKESAIIPFYIKGNKLVPCLSKEKISSL  100

G* *   *  *C**   **   * **  * ** * **  *       * *
GININNNDNTECVETSKAGISNISFEFSSLRLFIAVPKNLLSEIDKISSK  150

  * G* *********              ** * * ** * W**     *
DIDNGIHALFFNYQVNTRLANNKNRYDYISVSPNINYFSWRLRNLFEFNQ  200

           ****** *       * *G* ***  ***   * G* * **
NNDEKTWERNYTYLEKSFYDKKLNLVVGESYTNSNVYNNYSFTGISVSTD  250

  M        * * * **A *   * * *    *    V**G*F**
TDMYTPSEIDYTPEIHGVADSDSQIIVRQGNTIIINESVPAGPFSFPITN  300

*    * * * * *  *    * *    ****  **   * ** *
LMYTGGQLNVEITDIYGNKKQYTVNNSSLPVMRKAGLMVYNFISGKLTKK  350

       *  *   *G     *****  *  **  *  G**  *G  G*
NSEDGDFFTQGDINYGTHYNSTLFGGYQFSKNYFNLSTGIGTDLGFSGAW  400

 * ** *      *    *  *  *   * * * *    *  * ***
LLHVSRSNFKNKNGYNINLQQNTQLRPFNAGVNFDYAYRKKRYVELSDIG  450

.         ** * ****  *  * * *      Y*
WHGNLYNQLKNSFSLSLSKSLNKYGNFSLDYNKMKYWDNAYDSNSMSIRY  500

                                * ***P*      * *   **
FFKFMRAMITTNCSLNKYQSYEKKDKRFSINISLPLTKDYGHISSNYSFS  550
```

```
        *      ** **   *  * * *                                    *
NANTGTATSSVGLNGSFFNDARLNWNIQQNRTTRNNGYTDNTSYIATSYA  600

   * * * **              * G ***   ** *    * * ***
SPYGVFTGSYSGSNKYSSQFYSASGGIVLHSDGVAFTQKAGDTSALVRID  650

       *   * * *   G **   *  *  * * **   **    *
NISDIKIGNTPGVYTGYNGFALIPHLQPFKKNTILINDKGIPDGITLANI  700

   **** ****   *    *   ** *   **  ****  *
KKQVIPSRGAIVKVKFDAKKGNDILFKLTTKDGKTPPLGAIAHEKNGKQI  750

  ****   * *          * * *    C         *
NTGIVDDDGMLYMSGLSGTGIINVTWNGKVCSFPFSEKDISSKQLSVVNK  800

 C
QC          802  (Seq. #8)
```

But comparison with the protein from another strain, the sequence data shows the proteins have areas of high homology dispersed throughout, especially the first 410 residues. CssD has 4 cysteines (residues 91, 112, and two near the C-terminus) which are conserved in all ushers.

A region of dyad symmetry is present 347 base pairs into the CssD gene in both clones. The calculated free energy value of these structures is -7.2 kcal.

DNA flanking the CS6 genes. When compared with another strain, the DNA sequences of the two clones diverge immediately 1downstream of cssD and 96 bases upstream of cssA. The non-homologous flanking regions have homology with five distinct insertion sequences. The homologies include 3% to 32% of each insertion sequence but not entire insertion sequences. The homology of and Iso-IS1 in E8775 continues beyond the clones we have sequenced and may be a complete insertion sequences in the native plasmids.

It should be noted that minor variation in bases of the peptides does not destroy antigenicity. A protein having at least 60% homology with the CS6 A and B proteins identified herein having conservative substitution would be expected to have desirable properties.

As indicated previously, bacteria transformed with plasmids which express the CS6-A and CS6-B proteins may be administered by mouth. If the transformed bacteria are attenuated strains, they may be delivered live. It is also possible to administer killed bacteria. Carbonated beverages such as carbonated water are particularly useful as carriers which are inexpensive. When the bacteria are administered in a carrier wherein the pH is not over 7, an antacid may be given with the bacteria.

The CS6 A and CS6 B proteins may also be at least partially purified and administered by mouth by means usually used in the art to deliver antigens to the intestinal tract, including in protected forms such as liposomes, microcrystals, microdroplets, as microencapsulated formulations or as enterically coated capsules.